# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



| 1. AGENCY USE ONLY (*Leave blank*) | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | 03/15/99 | Final; 12/01/97 - 05/31/98 |

**4. TITLE AND SUBTITLE**

Inverse Protein Folding

**5. FUNDING NUMBERS**

N00014-92-J-1240


**6. AUTHOR(S)**

Ken A. Dill

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

University of California, San Francisco
3333 California Street, Suite 415
San Francisco, CA 94118


**8. PERFORMING ORGANIZATION REPORT NUMBER**


**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

Office of Naval Research, Code 341MCE
Molecular, Cellular, and Environmental Biology Program
800 North Quincy Street
Arlington, VA 22217-5680


**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

**11. SUPPLEMENTARY NOTES**

**12a. DISTRIBUTION/AVAILABILITY STATEMENT**

Approved for public release; distribution unlimited

**12b. DISTRIBUTION CODE**

**13. ABSTRACT** (*Maximum 200 words*)


We have developed Geocore, a computer algorithm that aims to predict the 3D structures of foldable polymers from their monomer sequences. Our first work was on foundations in simplified models, and most recently we have completed tests in which we found the method is successful in 18 peptides and small proteins. Our continuing efforts will be to speed it up and extend it to larger proteins.


19990323 042

**14. SUBJECT TERMS**

Protein folding, polymer design, computer algorithm, peptide modeling

**15. NUMBER OF PAGES** 3

**16. PRICE CODE**

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| Unclassified | Unclassified | Unclassified | UL |

GRANT #: N00014-92-J-1240

PRINCIPAL INVESTIGATOR: Dr. Ken A. Dill

INSTITUTION: University of California, San Francisco

GRANT TITLE: Inverse Protein Folding

AWARD PERIOD: 12/01/97 - 05/31/98

OBJECTIVE: To develop a model, called Geocore, for protein and polymer folding.

APPROACH: The model is united atom, with 3-6 discrete phi psi backbone options per amino acid and sidechains are fixed in their predominant rotameric state, chosen according to the relative populations in the Protein Databank. There is one hydrophobicity and one H bond parameter, steric violations are not allowed, and the conformations are searched by an extensive branch-and-bound method.

ACCOMPLISHMENTS: We used lattice models first to explore principles of fast conformational search processes and simple energy functions. We developed theory for inverse folding, on sequence design, that showed that this problem is in a much simpler class of computational difficulty than protein folding. We developed a very fast lattice-based search methodology that finds globally optimal conformations of the longest chains yet found by any method.

We then incorporated some of the same conformational search principles into an off-lattice method, called Geocore. We have tested Geocore now on 18 peptides - all the peptides having known structure (i.e., in the PDB) shorter than 31 amino acids (the limit of the current search method). Eleven of these peptides are water soluble. Geocore does well as a filtering algorithm, meaning that keeping the 400 lowest-energy configurations out of more than $10^9$ explored per typical peptide, is sufficient to retain a conformation that is native-like by the criteria that the secondary structures are similar and RMSD errors are less than about 5 angstroms, often about 3 angstroms.

The limitations of the method are in the limited phi psi options and the slow search speeds. We currently have independent funding to pursue improvements in these areas.

CONCLUSIONS: We have found that branch-and-bound search methods are competitive in speed with standard methods, like Monte Carlo, for models having discrete phi psi options. The advantage of the Geocore search is that we can be assured of global minima, and this gives a deep test of the energy function. A main conclusion from this work is that simple energy functions, with few parameters, are arguably no worse than more complex ones. We believe Geocore is a viable early-stage peptide folding model.

SIGNIFICANCE: This work takes steps toward a computer algorithm that can compute the structures of peptides and proteins, ab initio, from their monomer sequences.

PATENT INFORMATION: No patents.

AWARD INFORMATION: President, Biophysical Society, 1998-1999; First Hans Neurath Award, Protein Society, 1998; member BBCA Study Section, NIH.

PUBLICATIONS:


1. Yue, K. and K.A. Dill. (1992) Inverse Protein Folding Problem: Designing Polymer Sequences. *Proceedings of the National Academy of Sciences USA* **89**: 4163-4167.

2. Yue, K. and K.A. Dill. (1993) Sequence-Structure Relationships in Proteins and Copolymers. *Physical Review E* **48**: 2267-2278.

3. Yue, K., K.M. Fiebig, P.D. Thomas, H.S. Chan, E.I. Shakhnovich, and K.A. Dill. (1995) A Test of Lattice Protein Folding Algorithms. *Proceedings of the National Academy of Sciences USA* **92**: 325-329.

4. Yue, K. and K.A. Dill. (1995) Forces of Tertiary Structural Organization in Globular Proteins. *Proceedings of the National Academy of Sciences USA* **92**: 146-150.

5. Dill, K.A., S. Bromberg, K. Yue, K.M. Fiebig, D.P. Yee, P.D. Thomas, and H.S. Chan. (1995) Principles of Protein Folding - A Perspective From Simple Exact Models. *Protein Science* **4**: 561-602.

6. Dill, K.A., H.S. Chan, and K. Yue. (1995) The Protein Folding Problem: Searching Conformations of Compact Chain Molecules. *Macromolecular Symposia* **98**: 615-617.

7. Yue, K., and K.A. Dill. (1996) Folding Proteins With a Simple Energy Function and Extensive Conformational Searching. *Protein Science* **5**: 254-261.

8. Dill, K.A., and A.R. Fersht. (1996) Editorial overview: Folding and Binding. *Current Opinion in Structural Biology* **6**: 1-2.